Verification of NMME Extreme Seasonal Temperatures

Dan C. Collins
NOAA Climate Prediction Center

DATA

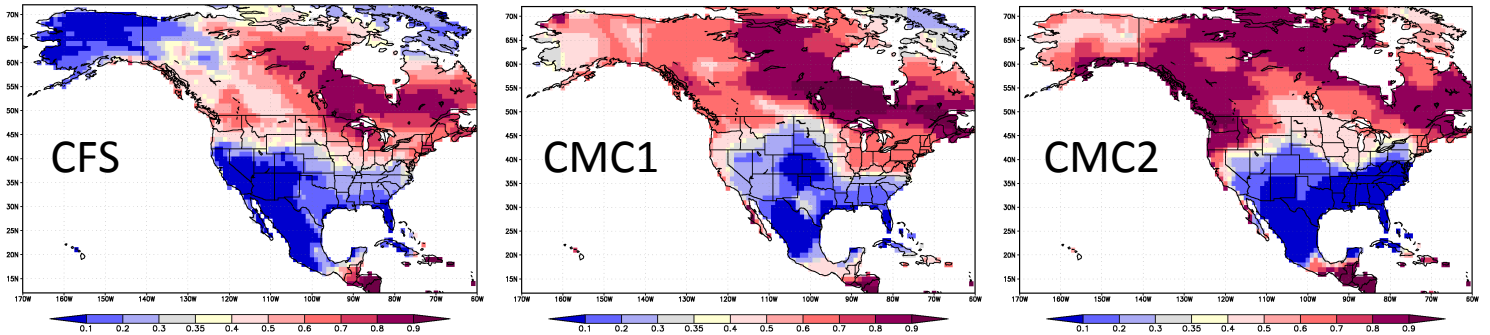
- NMME hindcast:
  - 12 x 1-month lead seasonal hindcasts

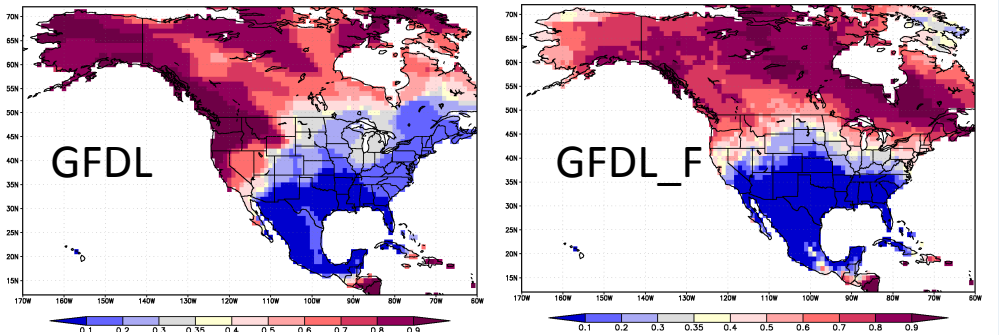
  (e.g., November 1st initialization predictions of DJF Temperature)
  - 29 year hindcasts, 1982 to 2010
- *GHCN+CAMS 2-m temperature observations*
- Considering the skill and calibration of probability forecasts for bottom and top deciles (i.e. 10^{th} and 90^{th} percentile forecasts)
- … and calibration of ensemble probabilities for above and below normal terciles simultaneously

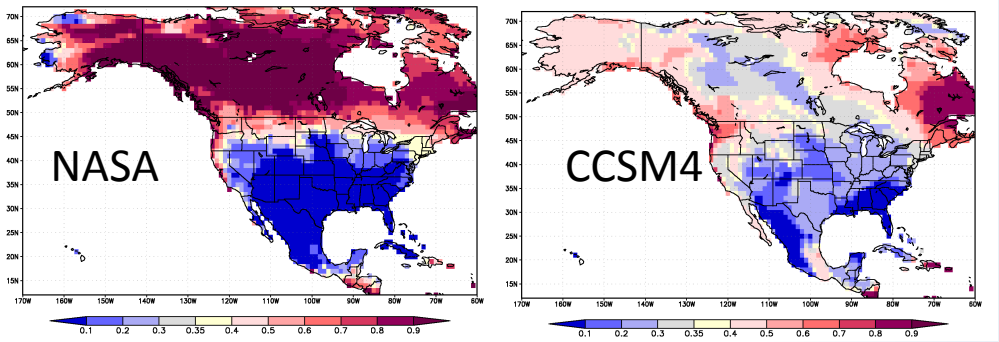
METHODS

1. **Count:** Probabilities based on percent count of ensemble members
   - Model climatological mean removed
   - Variance of members corrected
   - Terciles (+- 0.43 standard deviations) and deciles (+- 1.28 standard deviations) derived from normal distribution fit to hindcasts
   - Cross-validated (Leave 1-year out) calculation of model mean and standard deviations
2. **Ensemble Regression or EReg** probability calibration methodology (Unger et al 2009).
   - Removes systematic bias and calibrates probability based on hindcast correlation
   - Cross-validated regression (Leave 1-year out) calculation of all regression parameters and probabilities
   - Fit Gaussian distribution around each ensemble member

Multi-model ensemble **count** forecasts







Hindcasts of the 1997-98 DJF climate of North America

Ensemble Regression

1) Derive a regression equation for the least-squared error solution between the ensemble members and the observation, based on the ensemble mean and observations.

$$F^*_{(m)} = aF_{(m)} + b$$

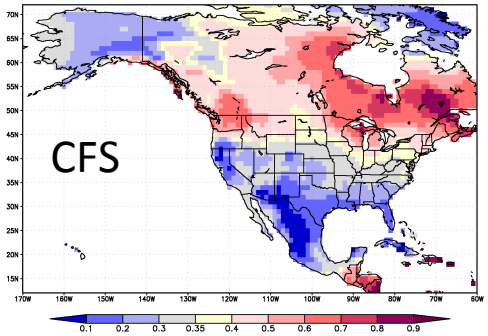
2) Expected residual error distribution of each member based on expected MSE of ensemble mean minus average ensemble spread.

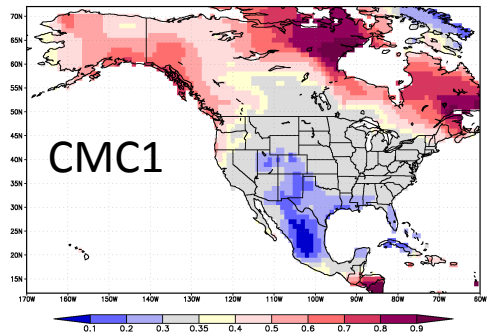
$$\sigma_\varepsilon^2 = [MSE] = \sigma_{ens}^2 + \varepsilon^2$$

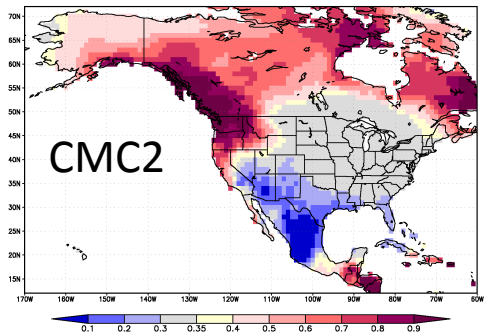
$$\varepsilon^2 = \sigma_\varepsilon^2 - \sigma_{ens}^2$$

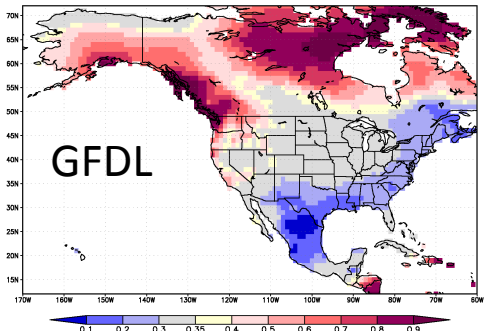
$$[MSE] = \sigma_{obs}^2 (1 - R_m^2)$$

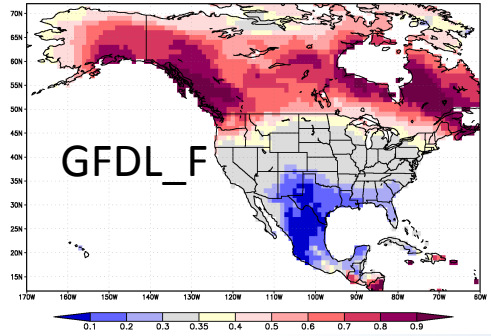
Multi-model ensemble **EReg** calibrated forecasts

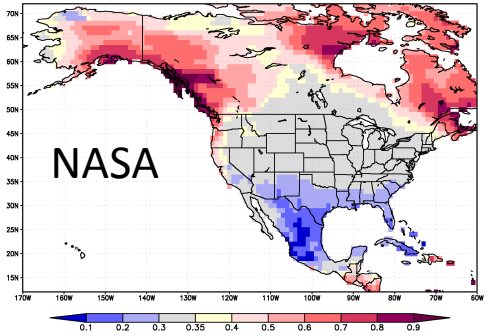


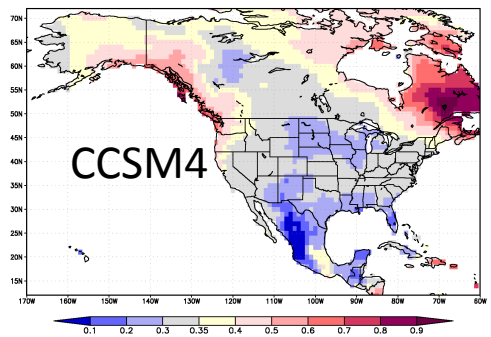










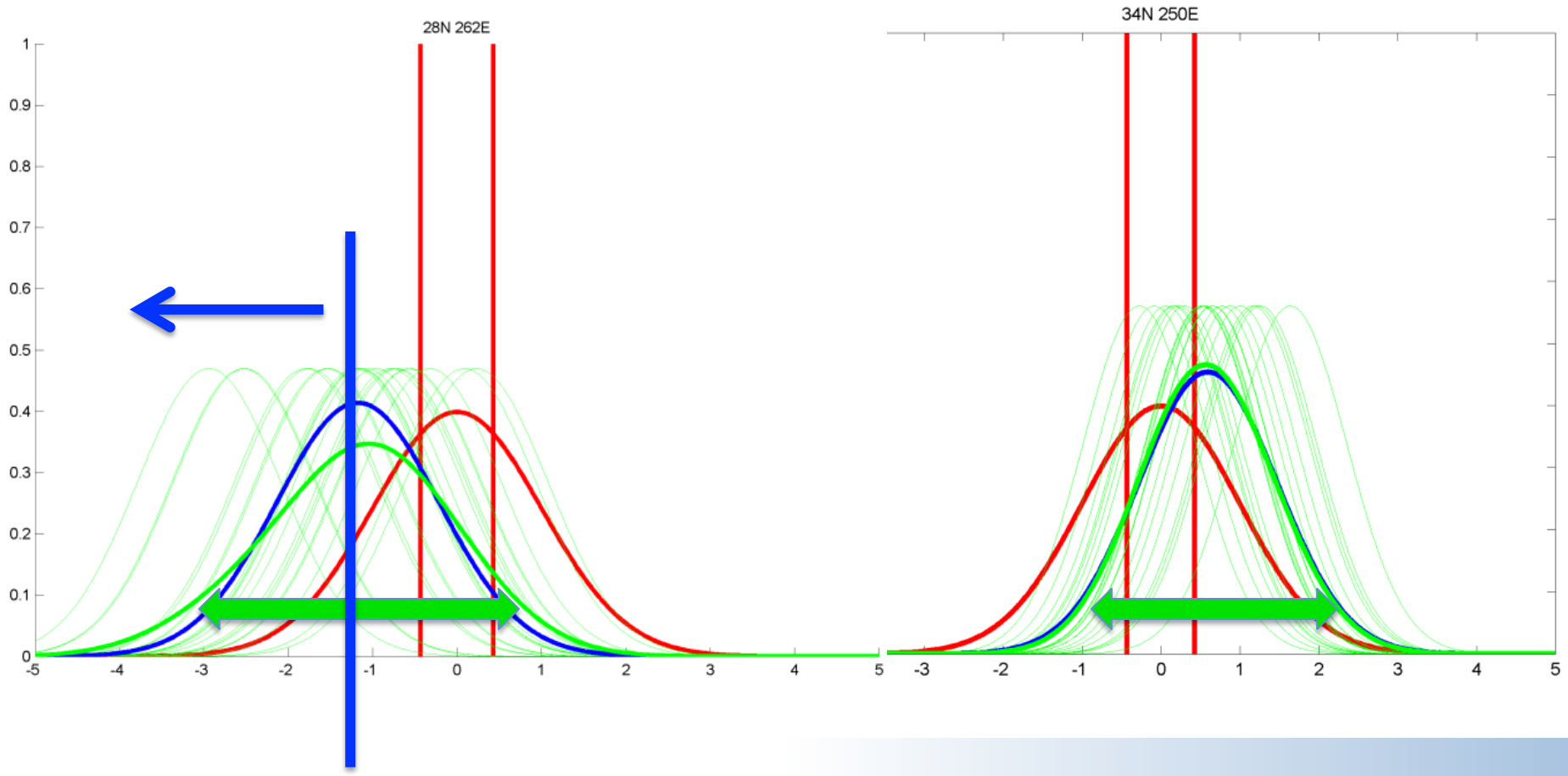


Calibration makes maps equally reliable and nearly equally skillful , but not the same.

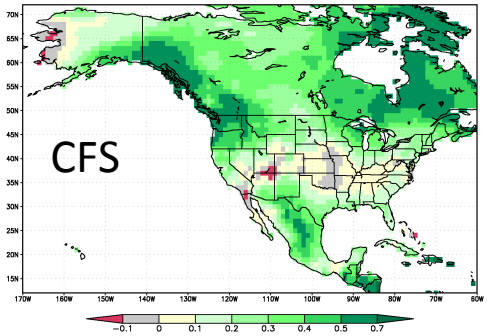
Seasonal Forecasting of Extremes & MME

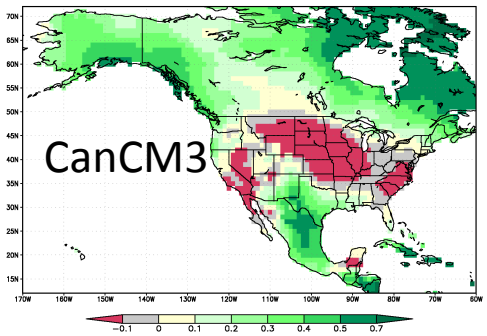
- Low predictability & extremes due to signal+noise
- Some individual models have 10 ensemble members ->
  - Poor resolution of tails of distribution (extremes)
- Probabilistic Outlooks for categorical forecasts
  - 3-category tercile forecasts (Above and Below normal)
  - Probability of extremes should be consistent with tercile forecasts
- Correlation used to adjust spread of model
  - Realistic probabilities representing hindcast skill
  - Improves **reliability**, while maintaining resolution of events
  - **Brier Skill Score** as metric (Resolution + Reliability)
- Multi-Model Ensembles
  - Many ensemble members reducing noise and canceling errors
  - **Better resolution of probability distribution, including tails (i.e. extremes)**

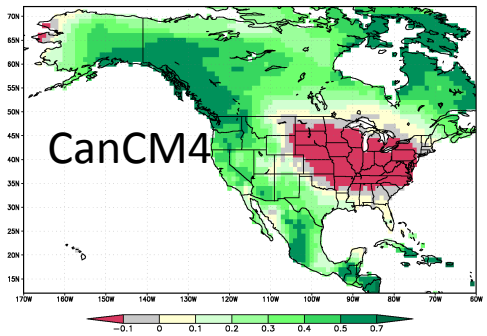
Variations in the mean and spread of ensemble members: Probability of extremes < -1.28 sigma, or > +1.28 sigma

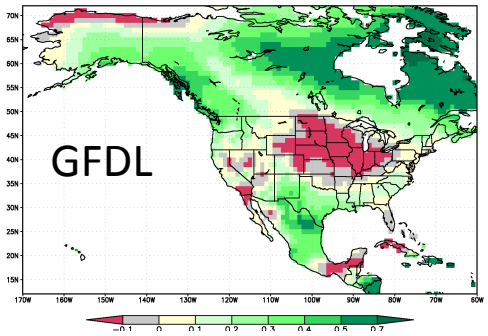


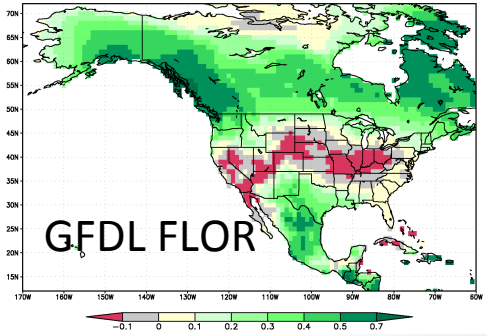
Cross-validated DJF Correlation

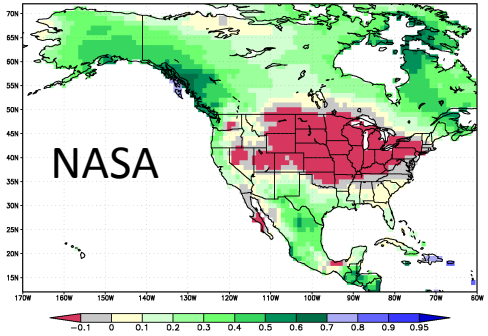


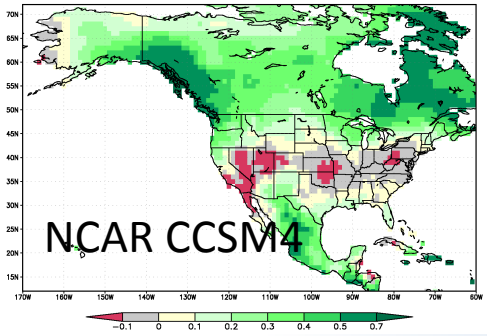


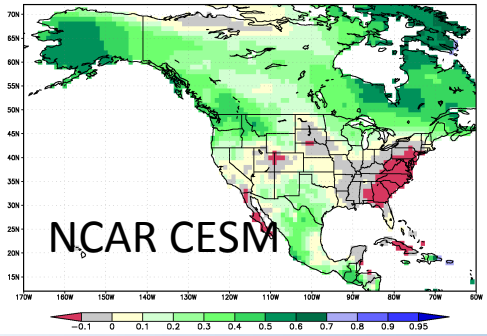




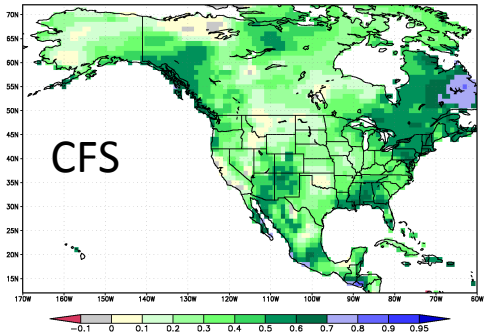


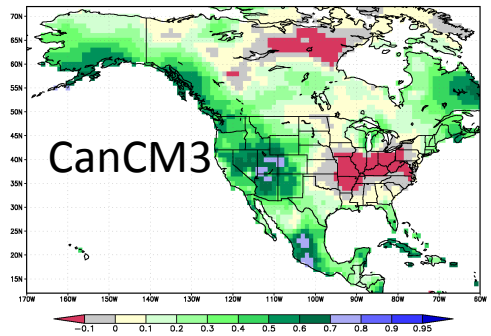


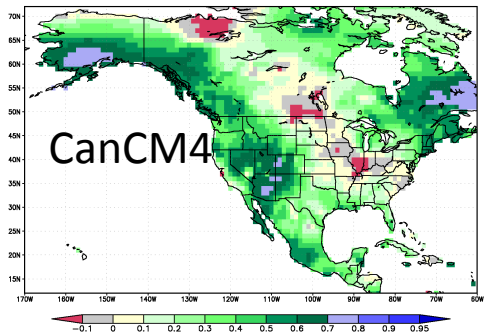


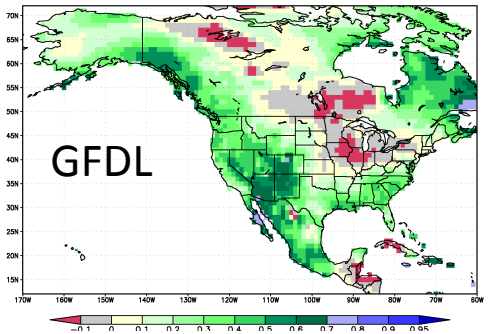


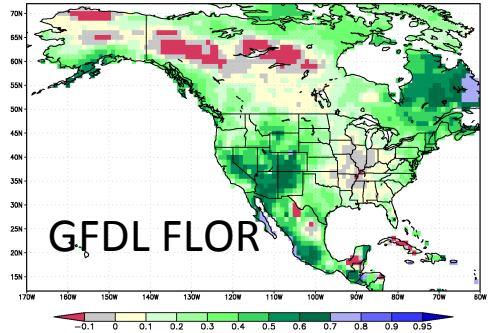
Cross-validated JJA Correlation

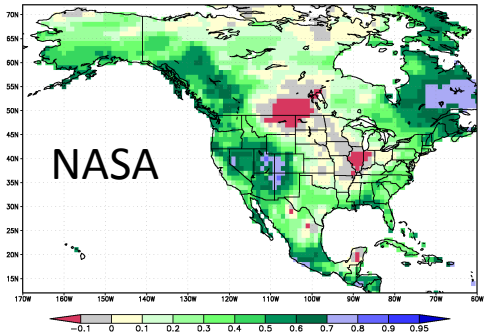


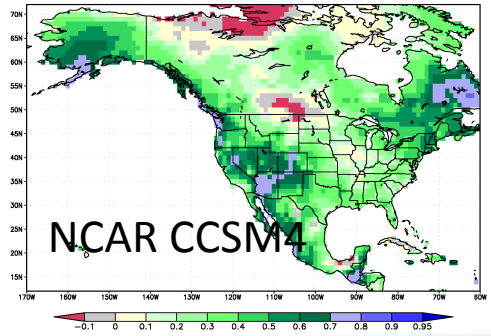


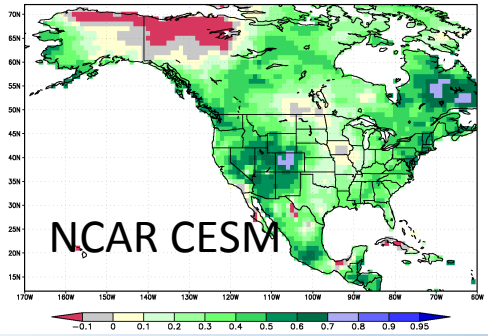




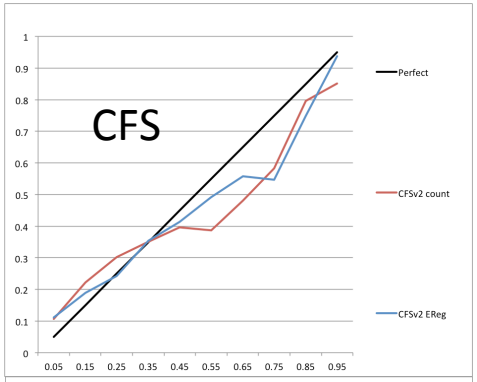


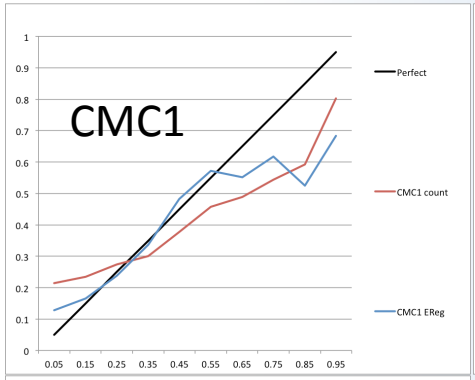


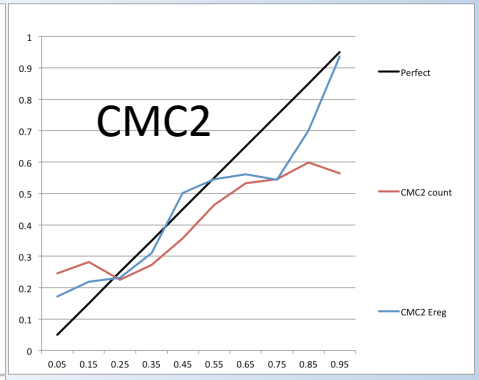


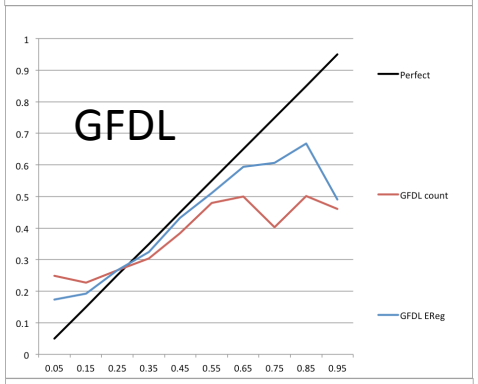


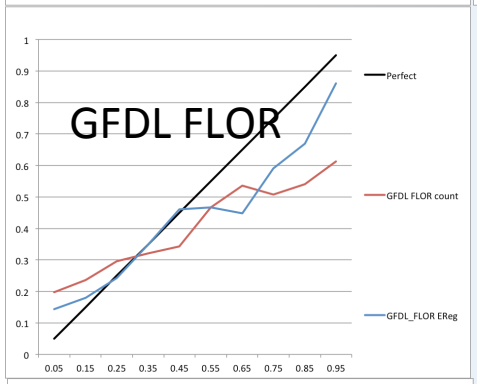
Each individual model's PDF is calibrated to improve reliability (DJF Terciles shown)

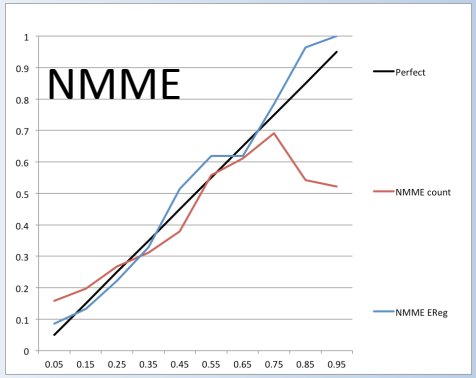


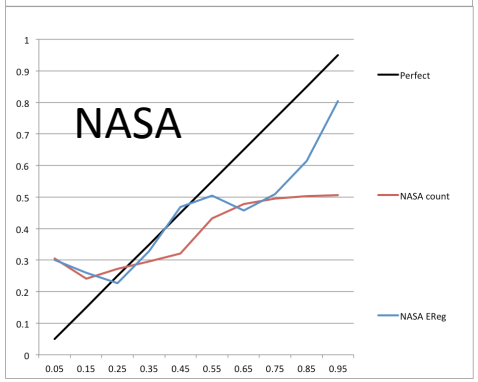


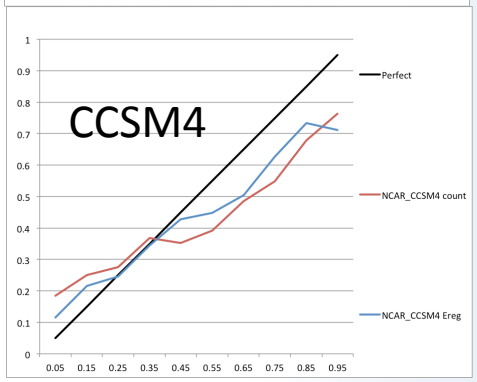






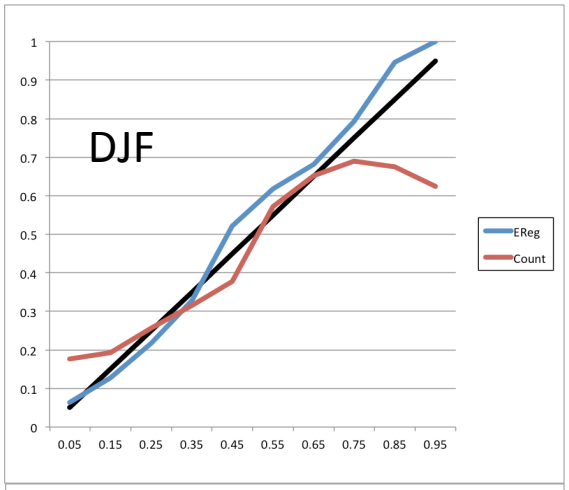


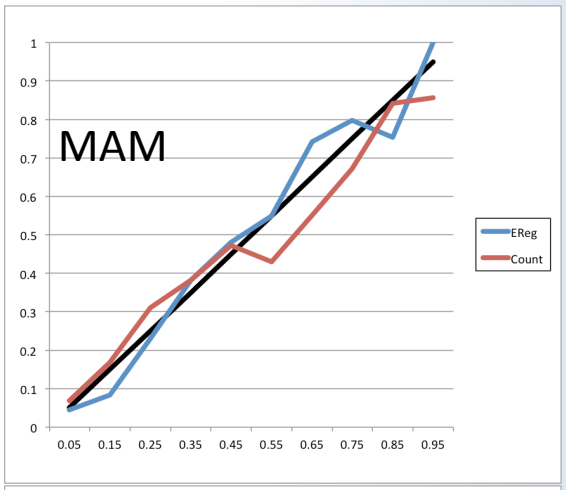


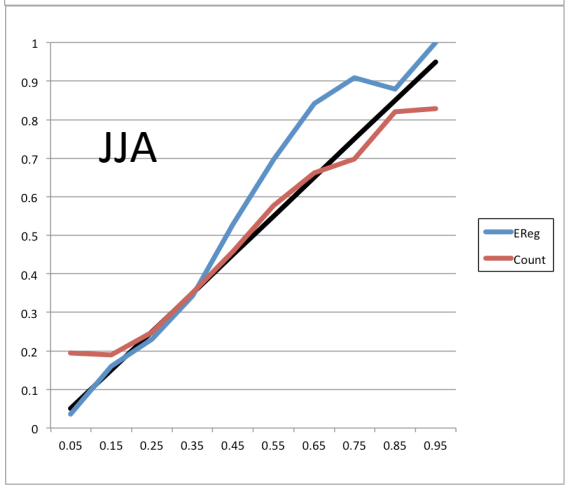


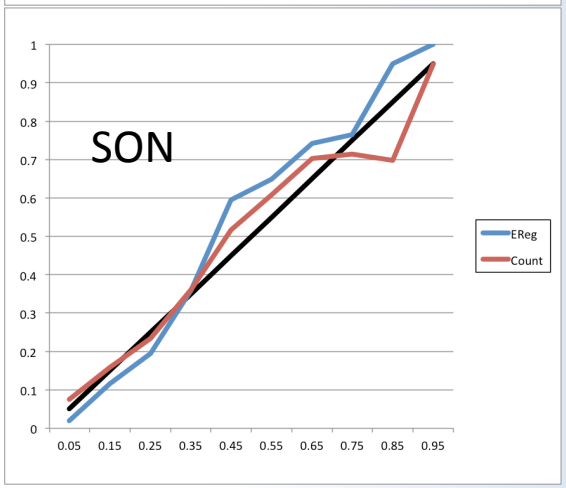
- EReg with all individual models -> improved reliability
- NMME improved both by MME combination & calibration
- ***Combined NMME is not adjusted*** for multi-model correlation

Combined NMME **Count** & **EReg** Tercile Probability Reliability



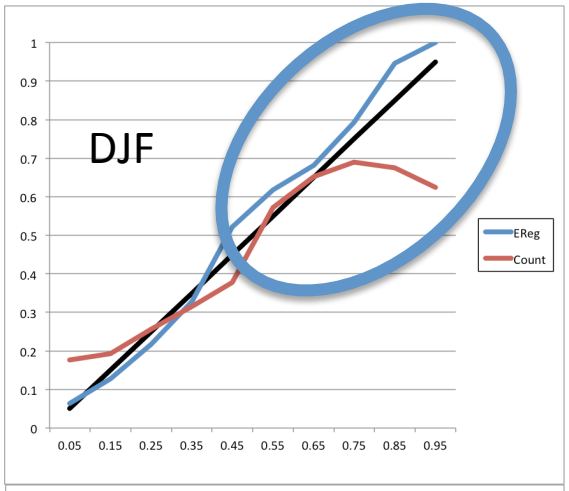


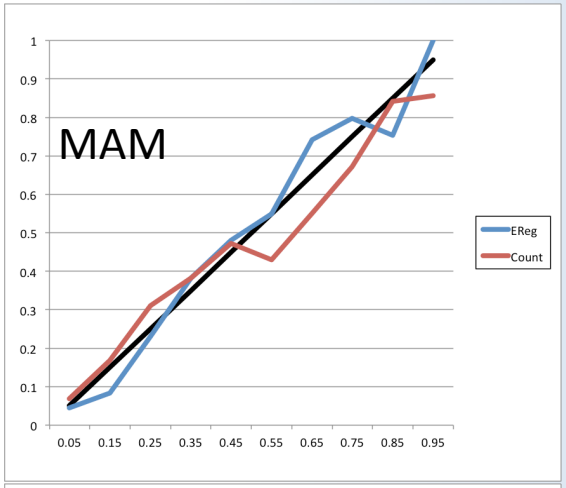


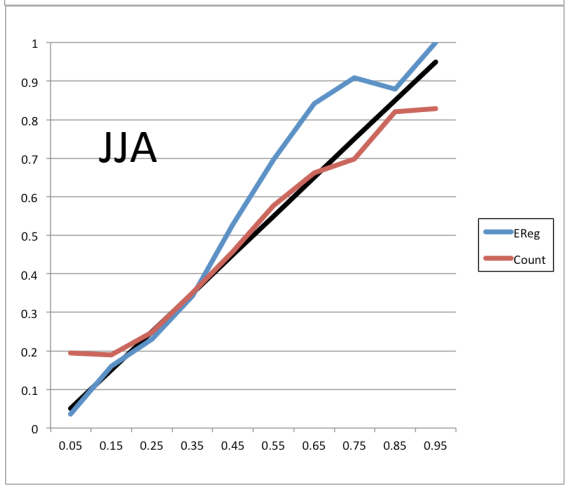


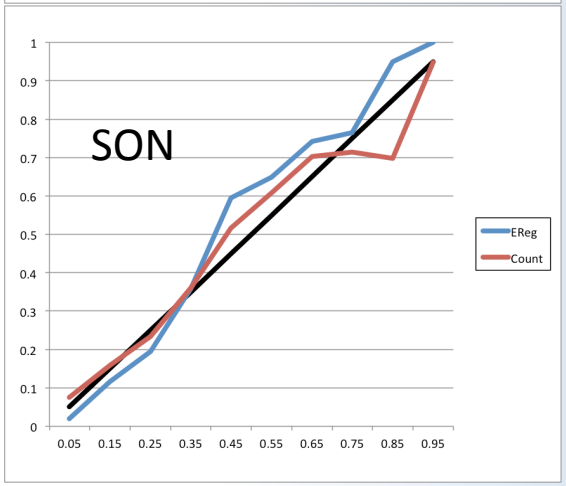
- Calibration improves DJF & MAM reliability
- Combined EReg NMME JJA & SON probabilities under-confident
- **Probabilities of combined NMME somewhat reliable. Combining calibrated models -> under-confident total probability**

Combined NMME **Count** & **EReg** Tercile Probability Reliability



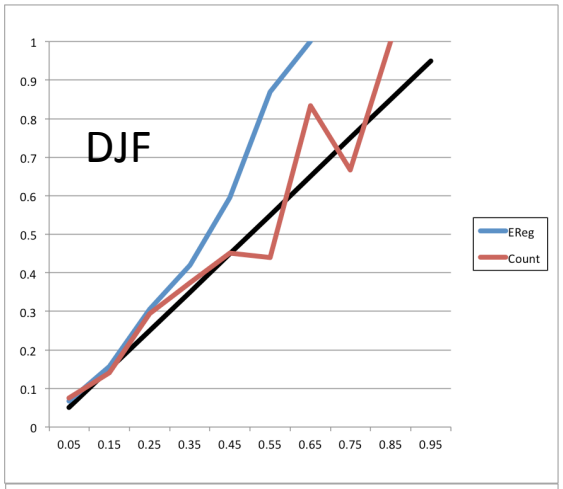


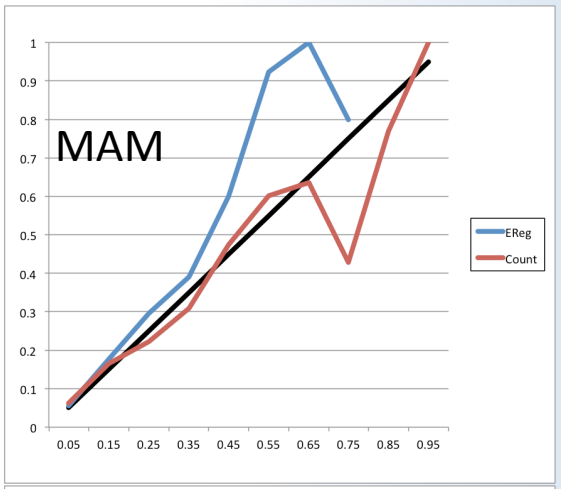


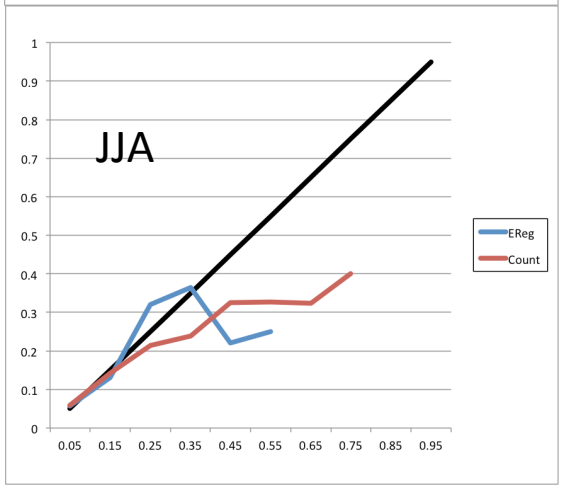


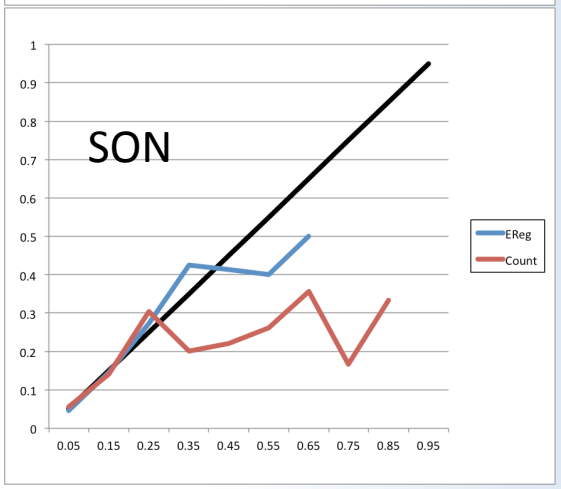
- Calibration improves DJF & MAM reliability
- Combined EReg NMME JJA & SON probabilities under-confident
- **Probabilities of combined NMME somewhat reliable. Combining calibrated models -> under-confident total probability**

Combined NMME **Count** & **EReg** Extreme Probability Reliability



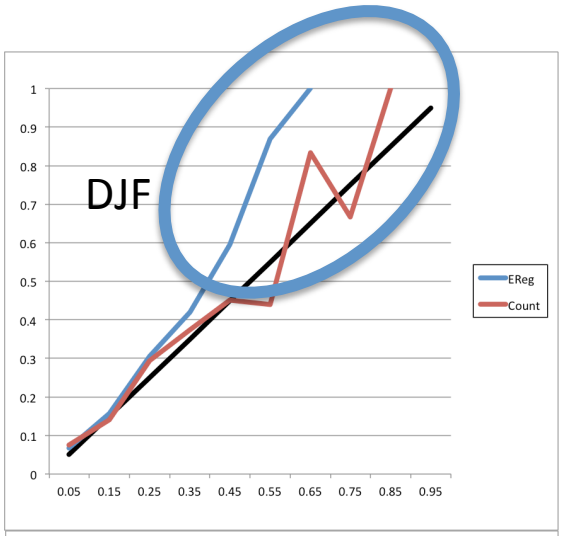


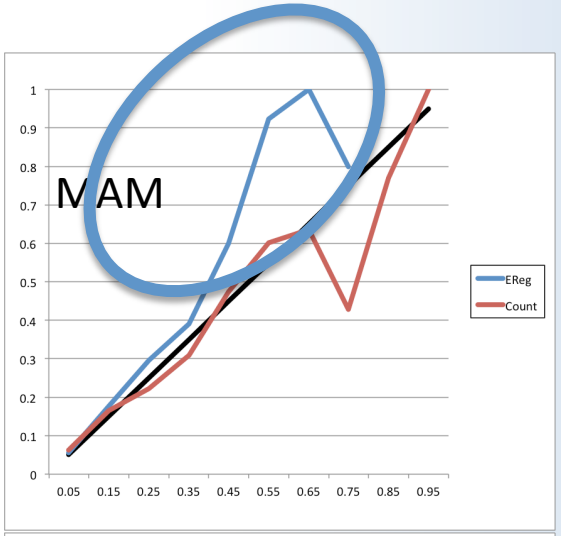


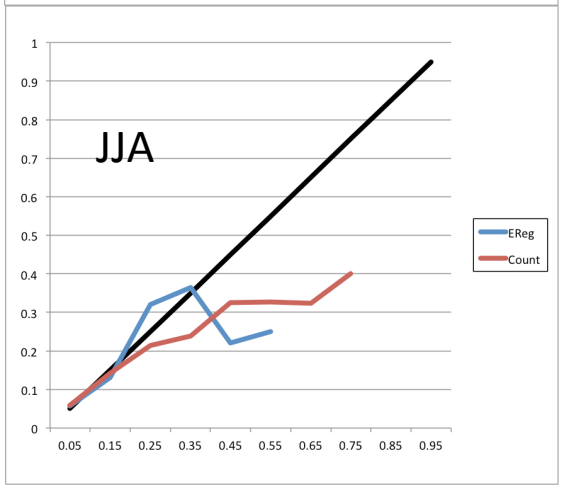


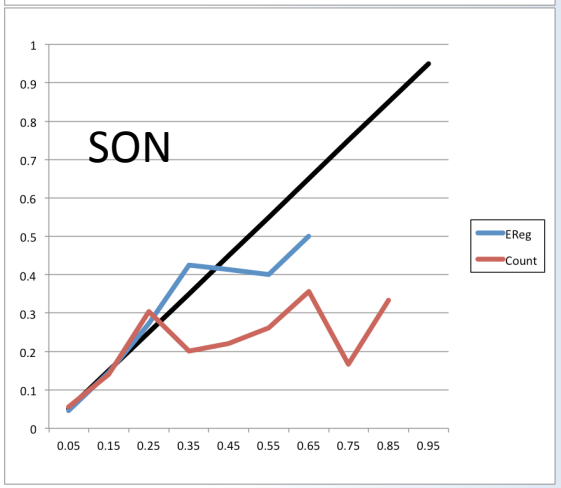
- Calibration of PDF produces under-confident forecasts in DJF & MAM
- Better reliability in some seasons

Combined NMME **Count** & **EReg** Extreme Probability Reliability



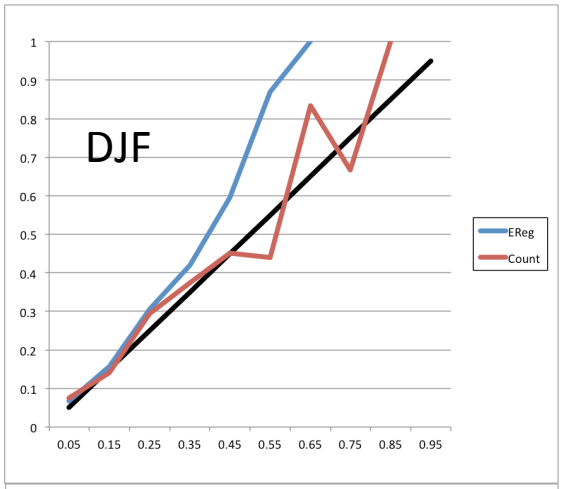


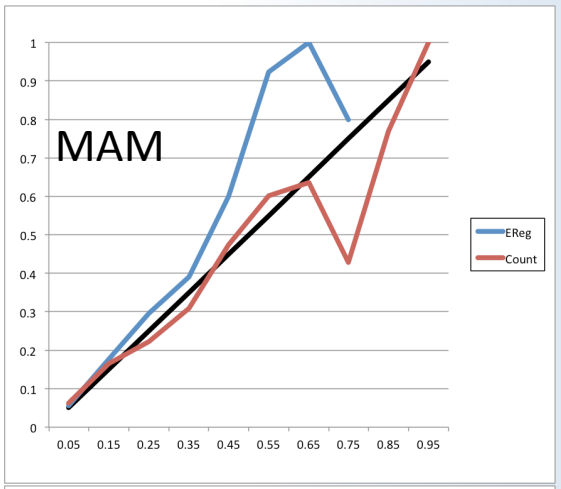


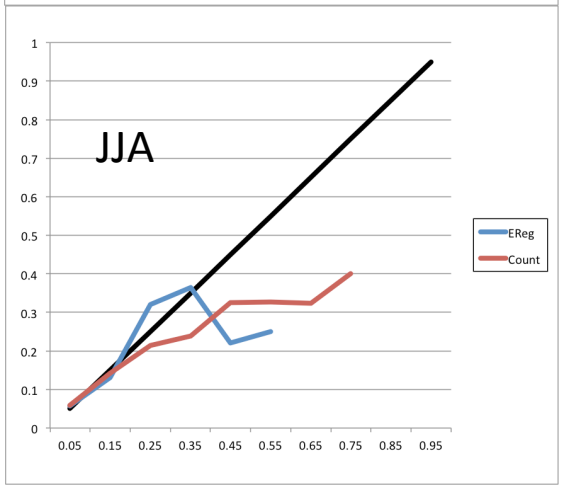


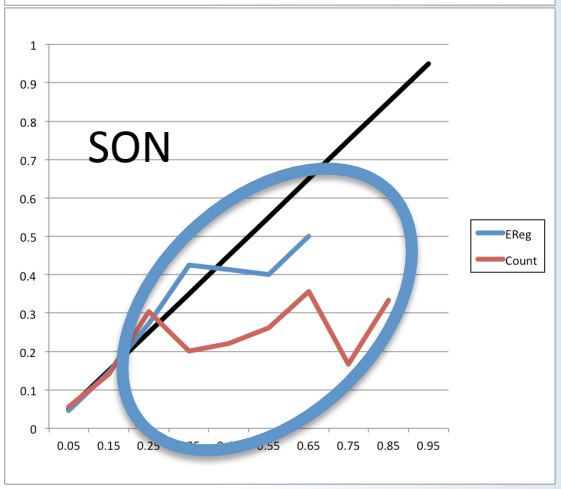
- Calibration of PDF produces under-confident forecasts in DJF & MAM
- Better reliability in some seasons

Combined NMME **Count** & **EReg** Extreme Probability Reliability



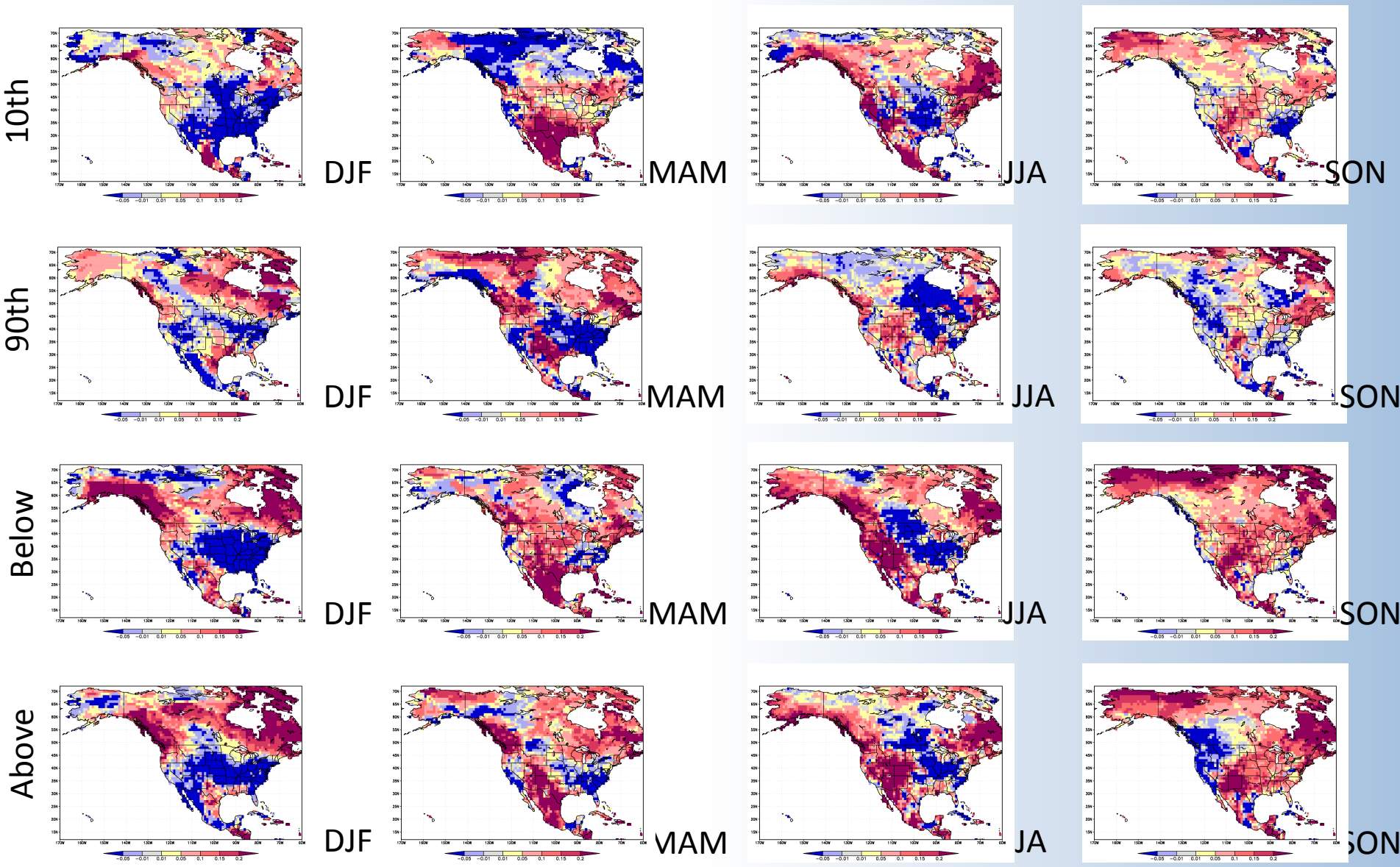




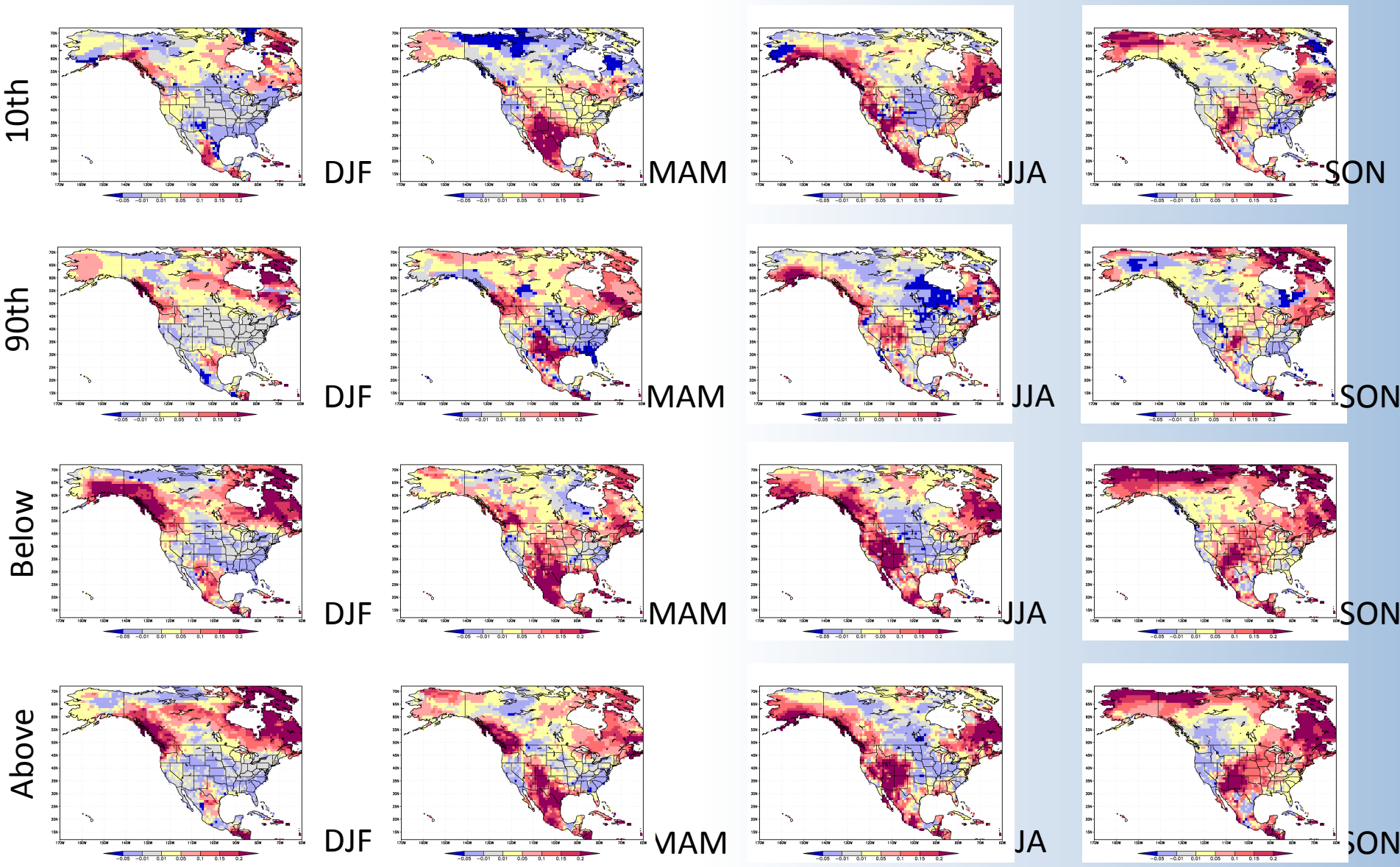


- Calibration of PDF produces under-confident forecasts in DJF & MAM
- Better reliability in some seasons

Four seasons 10th (top) and 90th (2nd row) percentile **count** hindcast Brier Skill Scores compared to below and above normal tercile BSS

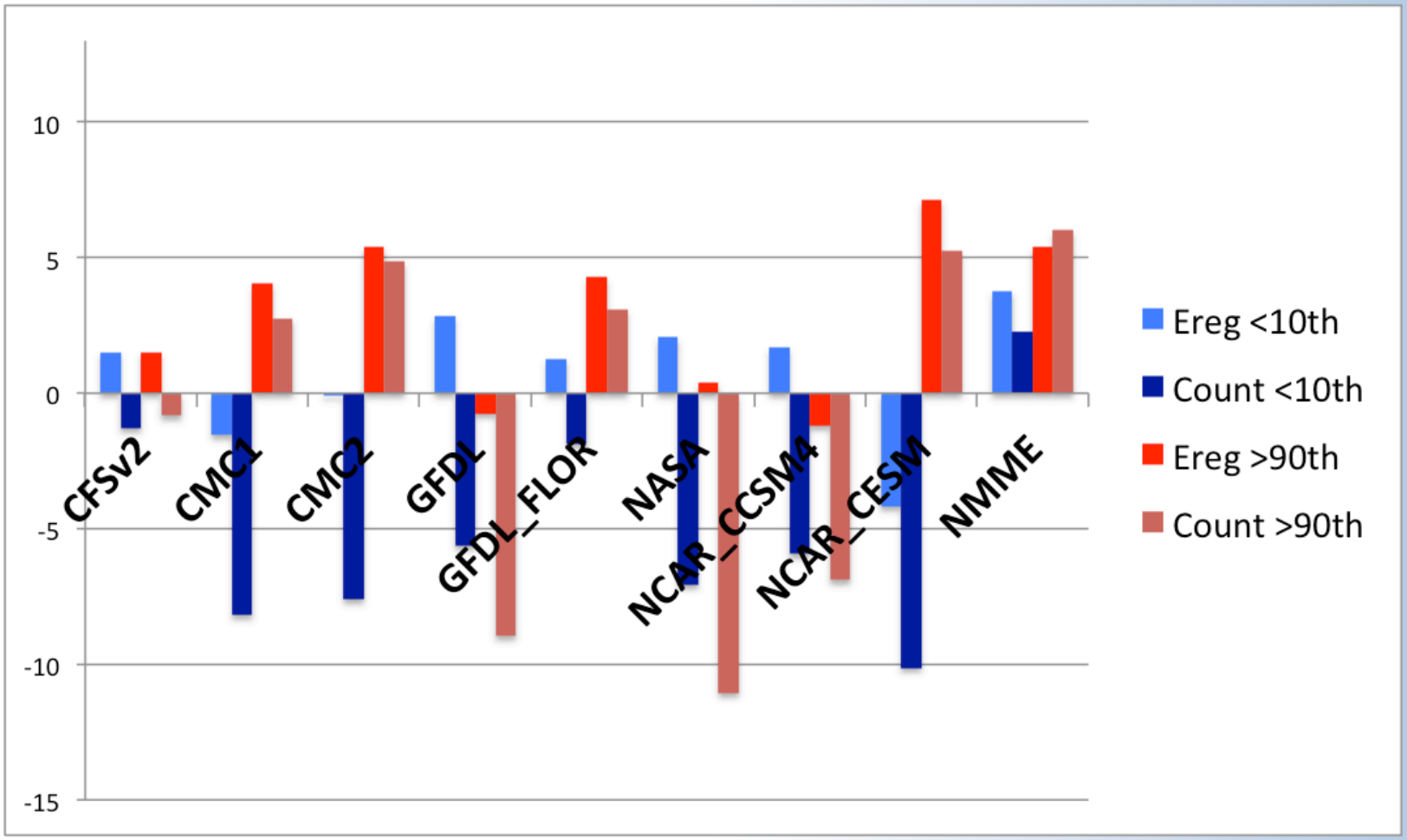


Four seasons 10th (top) and 90th (2nd row) percentile **EReg** hindcast Brier Skill Scores compared to below and above normal tercile BSS



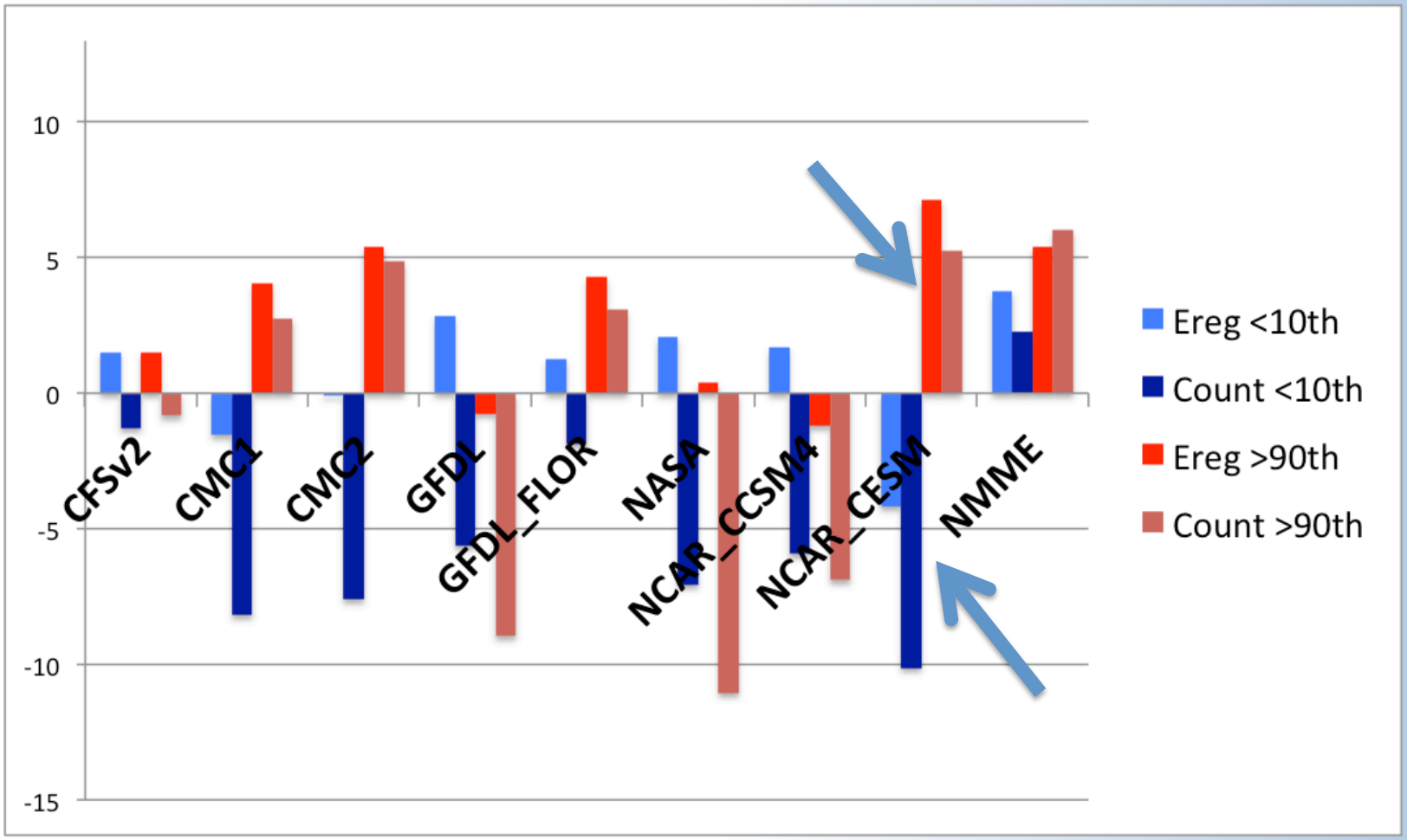
North America average Brier Skill Scores

DJF climatological 10th (blue) and 90th (red) percentile hindcasts
8 models and the combined NMME, bias corrected **Count** (bright) &
Ensemble Regression or **Ereg** calibrated (darker colors)



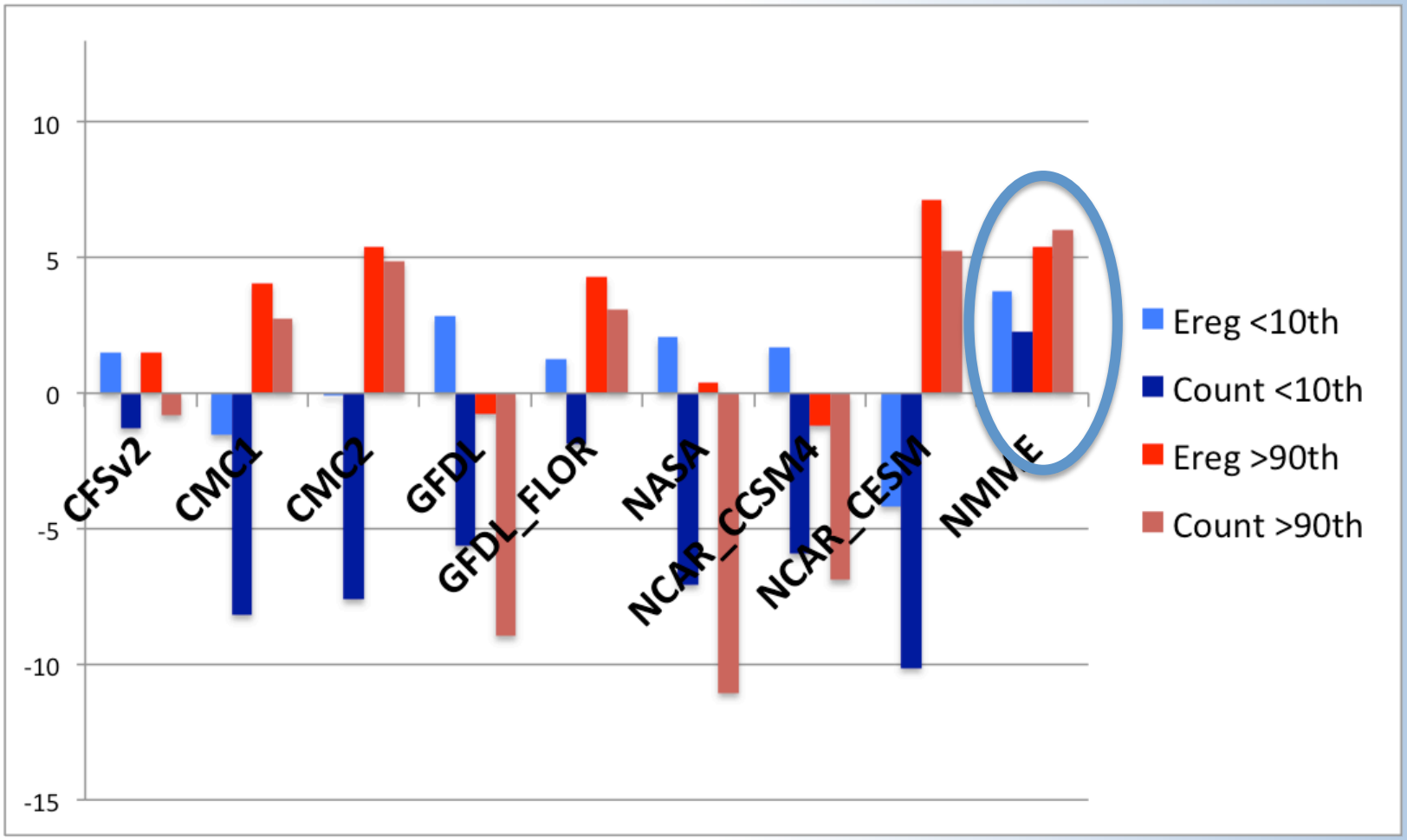
North America average Brier Skill Scores

DJF climatological 10th (blue) and 90th (red) percentile hindcasts
8 models and the combined NMME, bias corrected **Count** (bright) &
Ensemble Regression or **Ereg** calibrated (darker colors)



North America average Brier Skill Scores

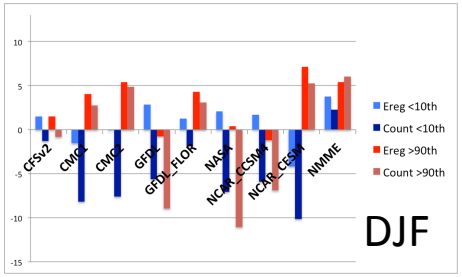
DJF climatological 10th (blue) and 90th (red) percentile hindcasts
8 models and the combined NMME, bias corrected **Count** (bright) &
Ensemble Regression or **Ereg** calibrated (darker colors)

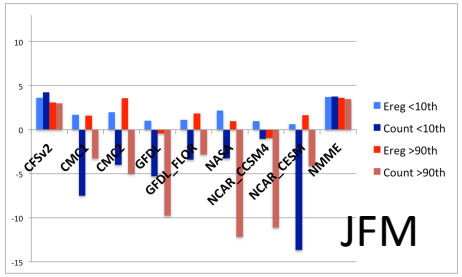


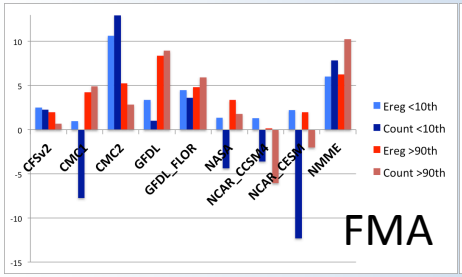
North America average Brier Skill Scores

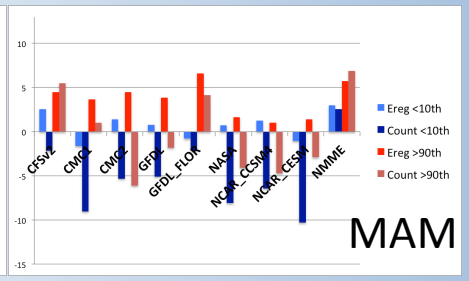
All 12 seasons 10th (blue) and 90th (red) percentile hindcasts

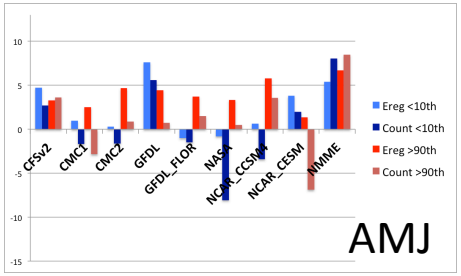
*Combined NMME (far right)

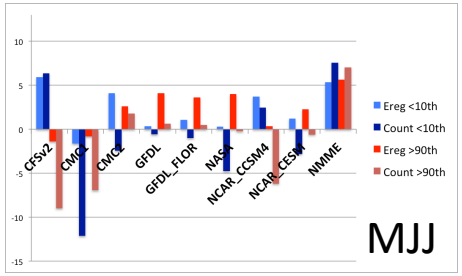


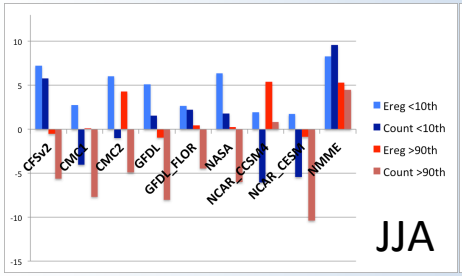


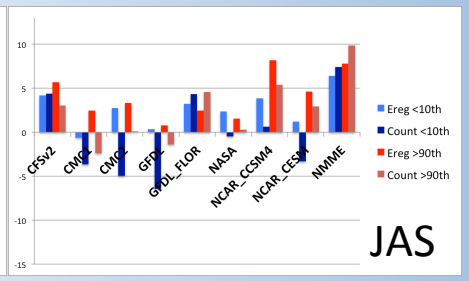


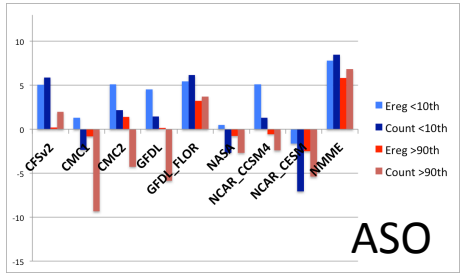


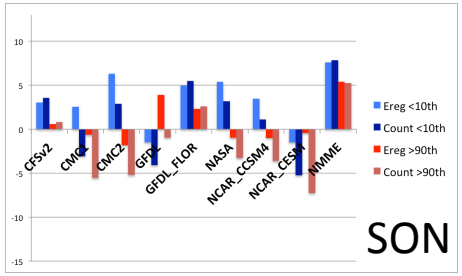


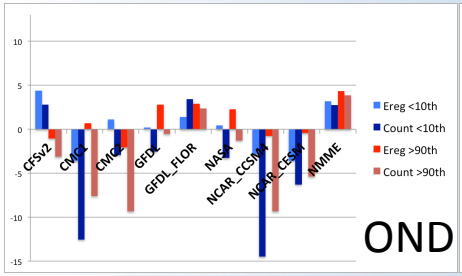


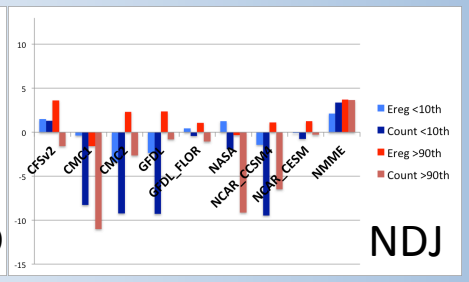








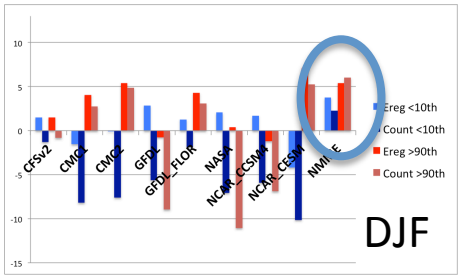


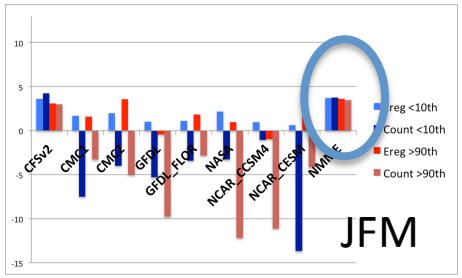


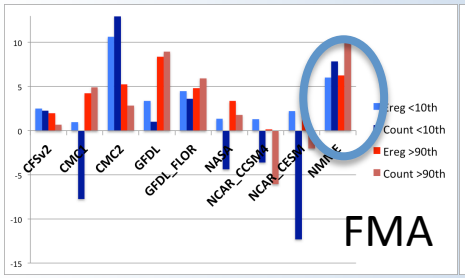
North America average Brier Skill Scores

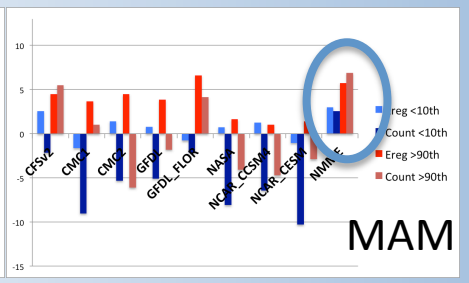
All 12 seasons 10th (blue) and 90th (red) percentile hindcasts

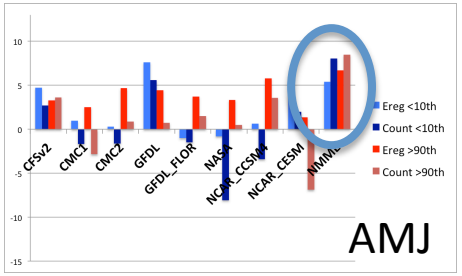
*Combined NMME (far right)

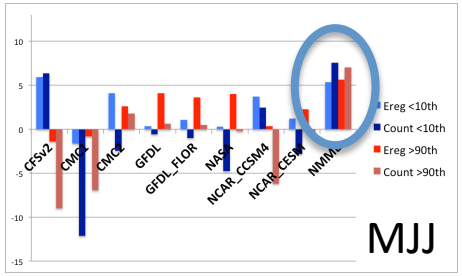


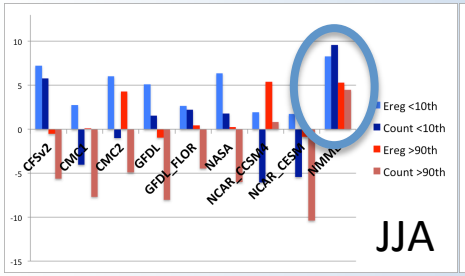


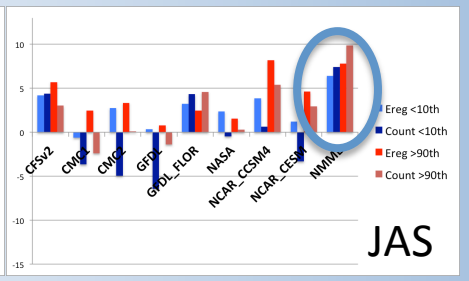


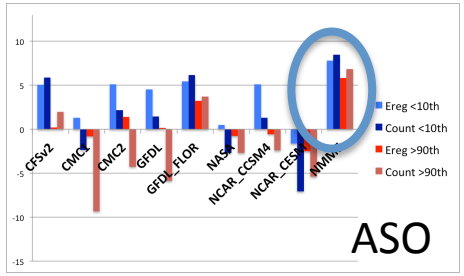


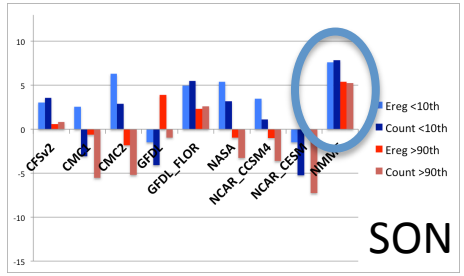


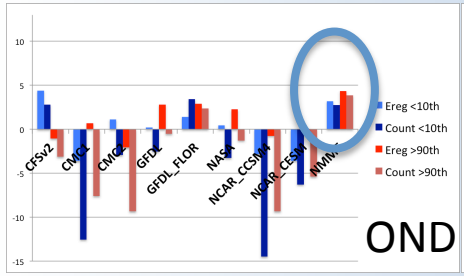


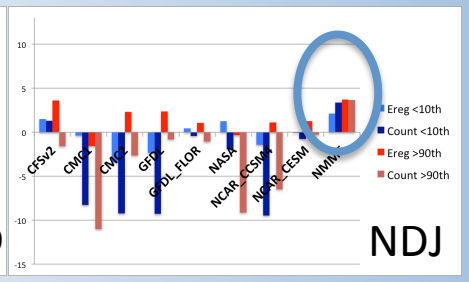












Combined NMME has positive skill in all seasons.

Not always the greatest skill, but consistently near the best model

Results & Conclusions

- Skill of tercile probabilities of seasonal forecasts implies skill of extremes
  - Skill in extreme forecasts in same regions as tercile forecasts

- Patterns of skill of 10^{th} and 90^{th} percentile forecasts similar; Differences appear related to decadal trends

- While individual ensemble models often have negative skill when forecasting extremes, on average over North America, **the combined NMME is found to have skill when forecasting extremes in all seasons**

- Regression calibration successfully removes areas of negative skill; however, combined NMME forecasts are sometimes under-confident

Thanks